Universidade Estadual de Campinas
Instituto de Computação

Marcos Vinícius Macêdo Borges

Semantic-enhanced recommendation of video lectures relying on ontology-based annotations

Anotação Semântica para Recomendação de Conteúdos Educacionais

CAMPINAS
2021

Marcos Vinícius Macêdo Borges

Semantic-enhanced recommendation of video lectures relying on ontology-based annotations

Anotação Semântica para Recomendação de Conteúdos Educacionais

Dissertação apresentada ao Instituto de Computação da Universidade Estadual de Campinas como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Dissertation presented to the Institute of Computing of the University of Campinas in partial fulfillment of the requirements for the degree of Master in Computer Science.

Supervisor/Orientador: Prof. Dr. Julio Cesar dos Reis

Este exemplar corresponde à versão final da Dissertação defendida por Marcos Vinícius Macêdo Borges e orientada pelo Prof. Dr. Julio Cesar dos Reis.

CAMPINAS
2021

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

B644s

Borges, Marcos Vinícius Macêdo, 1995-
Semantic-enhanced recommendation of video lectures relying on ontology-based annotations / Marcos Vinícius Macêdo Borges. – Campinas, SP : [s.n.], 2021.

Orientador: Julio Cesar dos Reis.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Computação.

1. Sistemas de recomendação (Filtragem da informação). 2. Ontologias (Recuperação de informação). 3. Aprendizado de máquina. 4. Web semântica. I. Reis, Julio Cesar dos, 1987-. II. Universidade Estadual de Campinas. Instituto de Computação. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Anotação semântica para recomendação de conteúdos educacionais
Palavras-chave em inglês:
Recommender systems (Information filtering)
Ontology (Information retrieval)
Machine learning
Semantic web
Área de concentração: Ciência da Computação
Titulação: Mestre em Ciência da Computação
Banca examinadora:
Julio Cesar dos Reis [Orientador]
Flavia Linhalis Arantes
Diego Addan Gonçalves
Data de defesa: 10-03-2021
Programa de Pós-Graduação: Ciência da Computação

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0001-7475-8883
- Currículo Lattes do autor: http://lattes.cnpq.br/2434058693586056

Universidade Estadual de Campinas
Instituto de Computação

INSTITUTO DE
COMPUTAÇÃO

Marcos Vinícius Macêdo Borges

Semantic-enhanced recommendation of video lectures relying on ontology-based annotations

Anotação Semântica para Recomendação de Conteúdos Educacionais

Banca Examinadora:

- Prof. Dr. Julio Cesar dos Reis
  Instituto de Computação, Universidade Estadual de Campinas (UNICAMP)
- Dra. Flavia Linhalis Arantes
  Núcleo de Informática Aplicada à Educação (NIED), Universidade Estadual de Campinas (UNICAMP)
- Dr. Diego Addan Gonçalves
  Instituto de Computação, Universidade Estadual de Campinas (UNICAMP)

A ata da defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

Campinas, 10 de março de 2021

Acknowledgements

First i would like to thanks Leticia, my girlfriend, for all the support, love and kindness to me, she helped me a lot in this process and could haven't finished this research without her.

Thanks to Julio, my supervisor, for his comprehension, his patience, and all the help during this master degree.

This work was financially supported in partial by the São Paulo Research Foundation (FAPESP) (grants #2017/02325-5 and #2018/00313-2)[1].

[1]The opinions expressed in here are not necessarily shared by the financial support agency.

Resumo

Sistemas de apoio à aprendizagem exploram diversos recursos multimídia para considerar individualidades do aluno bem com diferentes estilos de aprendizagem. Todavia, a crescente quantidade de conteúdos educacionais disponíveis em diferentes formatos e de maneira fragmentada dificulta o acesso e compreensão dos conceitos em estudo. Embora a literatura tenha proposto abordagens para explorar técnicas de recomendação que permitem representação explícita de semântica por meio de artefatos como ontologias, essa linha não foi totalmente explorada e ainda requer muitos esforços de pesquisa. Esta pesquisa objetiva conceber um método de recomendação de conteúdo educacional explorando o uso de anotações semânticas sobre transcrições textuais de videoaulas. As anotações servem como metadados que expressam o significado de trechos das aulas. A técnica de recomendação, como principal contribuição esperada, fundamenta-se nas anotações disponíveis para definir estratégias de ranking de conteúdos disponíveis a partir da proximidade semântica dos conceitos combinadas com técnicas de aprendizagem de máquina. A contribuição envolve o desenvolvimento de protótipos funcionais de software para validação experimental com base em conteúdos de videoaulas reais e deve destacar as principais vantagens e limitações da abordagem. Os resultados obtidos permitirão o acesso à recomendações mais adequadas para melhorar o processo de aprendizagem apresentando a possibilidade de uma experiência mais satisfatória pelos alunos.

Abstract

Learning support systems explore several audio-visual resources to consider individual needs and learning styles aiming to stimulate learning experiences. However, the large amount of online educational content in different formats, and the possibility of making them available in a fragmented way, turns difficult the tasks of accessing these resources and understanding the concepts under study. Although literature has proposed approaches to explore explicit semantic representation through artifacts such as ontologies in learning support systems, this research line still requires further investigation efforts. In this MS.c. dissertation, we propose a method for recommending educational content by exploring the use of semantic annotations over textual transcriptions from video lectures. Our investigation addresses the difficulties in extracting entities from natural language texts in subtitles of videos. Our work studies how to refine concepts in a domain ontology to support semantic annotation of video lecture subtitles. We report on the design and evaluation of a video lecture recommendation system which explores the extracted semantic annotations. In a case study, our solution explored semantically annotated videos with an ontology in the Computer Science domain. Obtained results indicate that our recommendation mechanism is suited to filter relevant video content in different use scenarios.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

1.1 Context and Motivation

Technology has drastically changed the way we live. Nowadays, we can easily exchange and share information. If we were not able to communicate so fast, we definitely would be living in a different society. In this context, the use of Information and Communication Technologies to support the provision of educational content is an irreversible trend. Technology can help to spread knowledge in many different ways. One of them is through video lectures. Teachers who have access to a camera can record his/her lecture and post it online. On the other side, people can watch recorded lectures as a student. Web has become an important tool to support the supplying of educational content. With the growth of information dissemination, Web has changed the availability of multimedia content that helps in the learning process, resulting in a significant increase in the amount of educational resources accessible to learners [35]. The number of online lectures available has grown very fast, which turns difficult for someone who wants to learn something new to chose which is the best video to watch.

In this context, efforts are required by students to select the appropriate resources in the learning process. Usually, several video lessons of a course are available for access. There are contexts in which students need to find similar online lectures that cover the same topic in a course. Potentially, video lessons from other courses or teachers can be interesting to replace or complement concepts under study of a lesson, or even to illustrate different examples. The challenge in this process is to find conceptually similar videos with a set of topics from a class. In addition, it is challenging for students to select and organize several related video lessons. It would be useful if people could obtain adequate recommendation from relevant video lectures.

Recommendation systems may serve as a solution to the problem of filtering most appropriate educational content in a personalized way. Recommendation systems are especially useful for users in online systems. In these systems, users witness a big amount of information that they are unable to handle. These systems play a key role in contexts in which decision making on the choice of items occurs in a large set of options [6].

1.2 Research Problem

This MS.c. Dissertation investigates how concepts and techniques concerning the Semantic Web [33] can allow new ways of recommending educational content from online video classes. Our proposal takes into account the meaning of concepts extracted from subtitles of video lectures.

In this context, the first challenge addressed in this MS.c Dissertation is to investigate techniques that allow semi-automatically annotating transcripts of video lessons based on Semantic Web standards. This must store structured and semantically enriched metadata. Annotation techniques explore the recognition of named entities in textual descriptions. A semantic annotation consists of associating a concept represented in an ontology [13] with a specific fragment of text or resource (*e.g.*, a video).

Ontologies [13] refer to syntactic structures that formally express the semantics of knowledge in a domain. Annotations allow the processed resources (*e.g.*, videos) to be better interpreted and combined with other available resources. In this context, the transcription of audios from video lectures in texts described in natural language can be an element used by recommendation systems. However, this entails several difficulties which require investigations as dealing with the quality of audio transcriptions, extracted subtitles and means of semantically exploring them. The use of techniques that purely compare texts suffers from problems inherent in syntactic processing of information. For example, polysemy problems and synonyms that appear frequently in texts make it difficult to reach high accuracy in existing techniques. In addition, an issue in this context refers to the quality and coverage of the underlying ontologies used for annotation purpose.

The second challenge tackled in this MS.c Dissertation is how exploring the semantic annotated resources for leveraging video lecture recommendation systems. Computational methods and tools capable of recommending content have been investigated in literature. However, meanings declared in online educational contents are still poorly studied for the aim of recommendation. Structures that represent knowledge and make explicit data semantics is the focus of the study of the Semantic Web. Obtaining means of associating semantics with online video lesson contents can support the development of more appropriate and accurate recommendation techniques. We understand that further investigations are necessary to better understand and achieve automatic techniques in this context.

1.3 Research Goals

This MS.c Dissertation aims to conceptualize and develop a recommendation method that explores semantic annotations generated from video lecture contents. We propose the use of formal and semi-formal ontologies as a means of representing knowledge about the content from video lectures. Our study provides the development of a software tool working as a video lecture player. As a result, learners can sign a doubt via the tool when watching a video. Our system recovers semantically related online lectures based on our defined recommendation mechanism based on semantic annotations encoding the videos. In particular, our obtained results are derived from the following specific objectives:

1. Study the refinement of concepts in a domain ontology to support semantic annotation of video lecture subtitles. This includes ways of adding new concepts in an existing ontology and adding new vocabulary description to enrich existing concepts in place;

2. Investigate and experimentally evaluate the use of semantic annotation techniques to generate semantic metadata from video lectures;

3. Investigate and evaluate the design of a video lecture recommendation system which explores the extracted semantic annotations;

1.4 Dissertation Organization

This MS.c. Dissertation is organized as a collection of two articles under review for publication. Each one of these articles corresponds to a chapter in this thesis, as follows:

Chapter 2 corresponds to the article "*Ontology Refinement for Supporting Semantic Metadata Extraction from Video Lecture Subtitles*", submitted to an international journal and is under review. This work proposed a technique for semi-automatically refining domain ontologies based on textual book summaries. Our proposal explores information extraction techniques and syntactical analyses for obtaining a refined version a given version of a domain ontology. We investigate to which extent our obtained refined ontology leverages semantic annotation tools to enable extraction of relevant semantic metadata from video lectures. We considered the domain of Computer Science for our conducted evaluations.

Chapter 3 corresponds to the article "*Video Lecture Recommendations based on Semantic Annotations*", submitted to an international journal and is under review. It presents our proposal as an architecture designed for our recommendation system. In our approach, we use the refined ontology and semantic annotation tools investigated in Chapter 2 for the generation of semantic metadata. This is the key input for the functioning of our recommendation technique. Our recommendation mechanism considers a content filtering in which videos are recommended based on their similarity regarding the semantic metadata (extracted based on our refined domain ontology) from video lecture subtitles. We present an implementation of the developed software tool and its evaluation in distinct scenarios.

Chapter 4 presents a in-depth discussion focused on each of the research objectives of this MS.c. Dissertation. Then, we discuss limitations and open perspectives reached from the study of our recommendation system.

Chapter 5 concludes this MS.c. Dissertation. It elaborates on how our research can be extended in future investigations. We highlight the overall contributions of this investigation and show how our research findings were disseminated.

Chapter 2

Ontology Refinement for Supporting Semantic Metadata Extraction from Video Lecture Subtitles

2.1 Introduction

Educational content through the web has become an important way for supporting the learning process, especially in virtual learning environments, resulting in a significant increase in the amount of educational resources available [35].

In this context, efforts are required by students to collect desired resources to help in their learning process. Usually, several video lectures of a course are available for access. There are situations where students need to find similar classes covering the same topic in a course. Potentially, video lectures from other courses or teachers may be interesting to replace or complement the concepts of a lesson, or even illustrate different examples. The filtering and searching of education contents could benefit from techniques exploring the meaning of concepts appearing in the video lectures.

The key challenge in this research is to investigate a technique that allow semi-automatically refinement of an ontology and evaluate its effectiveness in annotations of text transcriptions from video lectures. This work generates structured and semantically enriched data regarding video lectures. Annotation techniques explore the recognition of named entities in textual descriptions [34] using ontologies to associate concepts to resources. This may improve systems that use the generated annotation for several purposes, e.g. information retrieval. Concerning the use of subtitles as video transcriptions, the difficulty in considering automatic generated transcriptions is the ability to generate meaningful text that capture specific terms of the subject. In addition, subtitles carry relevant information about the moment where specific concepts are approached in a video lecture. It is essential for a system that relies on ontologies or taxonomies, the ability to understand key concepts from the application's domain. In this context, well defined knowledge sources improve this kind of systems. However, it is difficult to design refined ontologies without the presence of a domain specialist. In this sense,ontology learning techniques are useful to enable better construction of domain ontologies from unstructured texts.

This investigation addresses the challenge of semi-automatically enrich a domain ontology [11] considering textual book resources. We use the generated ontology in the task of creating annotations in text as metadata. Generated metadata associate ontology concepts with a particular piece of text or resource, such as a video or an audio transcription. Annotations allow better interpretation and combination of available resources. Many aspects demand careful research in semantic annotating videos using ontologies, including: (1) the use of automatic transcription to generate input text for metadata extraction from subtitles; (2) the effectiveness of these transcriptions considering the use of Portuguese language; and (3) the viability of the enriched ontology in this task and the possibility of using general-domain ontologies (*e.g.*, *DBpedia* ontology) in semantic annotations.

In this work, we show our designed technique for ontology enrichment using book summaries. We study to which extent the refined generated ontology leverages semantic annotation tools to enable extraction of relevant semantic metadata from video lectures. These metadata must be able to describe the video well, so it can be used as input to automatic semantic-enhanced recommendation methods. For our technique, we consider a set of Computer Science books and an initial computer science ontology to enrich its concepts.

For the semantic annotation, our study takes an *URL* of a video lecture from *YouTube* and automatically extracts the relevant semantic metadata from it. The semantic metadata refer to semantic annotations in natural language texts from the video subtitle content. Our experimental evaluation considers the investigation of the effects in using distinct ontologies. Our contribution enables further analysis of semantically annotated videos using existent annotation tools associated with general-domain or specific ontologies. We investigate how to improve annotation results by enriching a Computer Science related ontology using additional resources.

The remaining of this work is organized as follows: Section 2.2 describes the fundamental concepts underlying this investigation and reports on related work. Section 3.3.2 presents our proposed technique for ontology refinement from textual book resources. Section 2.4 presents the evaluation protocols for the experimental evaluation conducted. Section 2.4.2 shows the obtained results in the context of semantic annotations for video lectures. Section 2.5 discusses the findings and challenges encountered in the conduction of this research; finally, Section 2.6 presents the conclusion remarks.

2.2 Background

Gruber [13] defines an ontology in the Computer Science context as a formal, explicit and shared specification of a conceptualization. The construction of ontologies requires the use of standardized structures to describe concepts in a given domain. Ensuring that ontologies present a structure to allow the reuse of information as well as that entities are well defined, it is necessary to formalize the representation of a vocabulary to be interpretable by machines. The W3C Web Ontology Language[1] (OWL) is usually adopted

[1]https://www.w3.org/2001/sw/wiki/OWL

to describe Web ontologies.

2.2.1 Ontology Learning

Ontologies can be created by extracting relevant information from text using a process called ontology population. However, handcrafting such big ontologies is a difficult task. Therefore, instead of handcrafting ontologies, a research trend is now shifting toward automatic ontology learning. Whenever an author writes something in the form of text, s(he) is actually doing it by following a domain model in his mind. (S)he knows the meanings behind various concepts of a particular domain, and then using that model, s(he) transfers some of that domain information to the text, both implicitly and explicitly.

Ontology learning is a reverse process as domain model is reconstructed from input text by exploiting the formal structure saved in author's mind [3]. The tasks of ontology learning can be decomposed in several aspects describing in which way the ontology is enriched: terms, concepts, relations and axioms.

The main approaches for term extraction from text in ontology learning rely mostly on sentence-tagging and statistical approaches. Hippisley *et al.* [15] used syntactic analysis to identify and extract complex terms in which complex terms take the role of hypernym word. The use of seed words is another methodology applied to guide ontology learning tasks [3]. Seed words are domain-specific words providing a base for other algorithms to extract similar domain specific terms and concepts. This technique ensures that only those terms that are more relevant and semantically closer to seed words are extracted. Sanchez and Moreno *et al.* [30] explored seed words to extract domain-specific documents from the Web and used them as input to extract terms and concepts for ontology construction.

In relation extraction, most of the techniques rely on the use of external sources of data and linguistic techniques to map terms into relations in ontology learning. Ciaramita *et al.* [8] applied dependency paths information presented in parse trees to find relationship patterns. For instance, for two specific concepts, new relations are found by extracting the shortest path among those concepts in a parsing tree. Lexicon-syntactic pattern analysis plays a key role in taxonomic and non-taxonomic relation extraction phases of ontology learning. To extract relations, this algorithm uses regular expressions in the sentences. For example, "NN is defined as NP... , NP" where NN is a Noun and NP is a Noun phrase. it is a sentence that can be used as a rule that extract patterns of hypernyms. This type of rule-based approach is useful in extracting *is–a* relationship or *part-of* relationship [19].

Other common approach used in several NLP tasks is the TF-IDF technique TF-IDF is a technique in Data mining field for measuring the word importance among a set of documents. Typically, topics are identified by finding the special words that characterize documents about that topic. There are certain words that appear rarely in a collection of documents, yet do not tell anything useful. On the other hand, there are words equally rare, but describe something about the document. The difference between rare words that tell us something and those that do not has to do with the concentration of the useful words in just a few documents [26]. The formal measure of how concentrated into relatively few documents are the occurrences of a given word is called TF-IDF (Term Frequency times Inverse Document Frequency). In the context of Ontology learning, it is useful where we have datasets as unstructured texts. In this context, the technique allows us getting the most relevant words in that corpus, so it can be used as input for other tasks, such as classification of documents or detection of candidates words to be concepts in ontologies.

2.2.2 Semantic Annotation

Semantic annotation refers to the process of creating semantic descriptors (metadata) about resources like texts described in natural language, images and videos. This process assigns a relationship between elements of the text and concepts described in an ontology [34]. A semantic annotation requires the use of an ontology offered as input for the semantic annotation task. In this context, the major difficulty is to achieve a set of annotations suited to identify the most important terms, capturing the key concepts from free-texts in terms of metadata.

One of the ways to produce semantic annotations from free-text content is through methods such as *Named Entity Recognition* (NER) and *Entity Linking* (EL) [24]. *Named Entity Recognition* refers to the process of recognizing parts of a text that may require some knowledge of the context by the reader to be better understood, such as names that refer to a person or a place. These parts are called entities [27]. From the entities recognized in a text, an *Entity Linking* method is responsible for establishing a link for additional content that exactly describe what the entity stands for. In the process carried out by an *Entity Linking*, we consider the set of entities represented as ontologies. In this work, we consider the use of semantic annotations to create metadata from video lectures transcriptions.

Our investigation identified software tools for semantically annotating texts such as *AutoMeta*[2], *CSO-Classifier*[3] [28], *NCBO Annotator* [23] and OntoText [17].

AutoMeta (Automatic Metadata annotation) refers to an environment for (semi)-automatic annotation and meta-annotation of documents for publishing on the Web using RDFa, a W3C recommended annotation language based on the RDF.

CSO-Classifier aims to classify content from scientific papers with the topics of the Computer Science Ontology (CSO) [28]. The proposal is to synthesize the content of papers to allow performing different kinds of analytic such as trend analysis, recommendation, find authors' topics of interest and topic analysis.

The National Center for Biomedical Ontology Annotator (NCBO) [23] is an ontology-based web service for annotation of textual biomedical data with biomedical ontology concepts. The biomedical community can use the Annotator service to tag datasets automatically with concepts from more than 200 ontologies coming from key biomedical ontology repositories.

OntoText [17] is a tagging service to enrich content by pasting an URL or a piece of text. It is based on data from *DBpedia* and *WikiData*. This tool explores machine learning algorithms to recognize mentions of entities such as Person, Organization, Location, and relationships between them. Our study investigated the effectiveness of these techniques

[2]https://github.com/celsowm/AutoMeta
[3]https://github.com/angelosalatino/cso-classifier

to annotate automatically created audio transcriptions from video lectures. Considering the informality of textual transcriptions and the use of the Portuguese language in videos, there is a lack of studies on the effectiveness of semantic annotations in such context.

2.2.3 Related work

Some studies have focused on semantic metadata extraction techniques relying in multimedia content as input. Santos *et al.* [31] defined an environment for the extraction of semantic metadata from soccer games videos. It used semantic annotation and ontologies to perform the transcription and automatically classification of videos.

Similarly, Coelho *et al.* [9] defined a framework to index video lectures based on semantic annotations of video's transcription using ontologies for describing the annotated terms. In their work, semantic annotation is used to create tags denoting the main topics covered in the video. These tags are used to search video lectures. Both investigations [31] [9] used the DBpedia ontology[4] as a basis to accomplish the semantic annotation.

Saraiva & Medeiros [32] defined a framework exploring taxonomy and graph oriented databases to correlate educational contents. In their work, different types of educational resources are considered. The concepts of each material are generally extracted and relationships are created between the different educational materials.

These studies have presented the potentialities of semantic annotation for the search of video lectures. However, the annotation process still remained in a high level, in which only the main topics of videos are considered. In our investigation, we provide fine-grained annotations based on subtitles in natural language text information. We investigate to which extent existing semantic annotation tools are suited to provide the necessary metadata. In addition, we study how our refined ontology can improve the annotation results.

2.3 OntoRef: Ontology Refinement through Book Resources

Our technique (cf. Figure 2.1) was designed for considering few steps to catch relevant terms to the domain without the need of a domain specialist. We aim to ensure that these terms are as correct as possible. Our technique is organized in the following steps: pre-processing of input sources; extraction of relevant terms from textual book resources; matching of terms to generate candidate ontology concepts. After this initial process, our technique considers a lexicon syntactic analysis where not only relevant terms, but their context in sentences is detected. This aims to reveal relationship between other concepts in the ontology and the candidate terms. At the final step, the BabelNet is taken and verified for enrichment of ontology concepts.

Pre-Processing. This step extracts content from text book files in *PDF* as unstructured text (raw text). Then all unstructured texts are Lemmatized and *stopwords* are removed.

[4]https://wiki.DBpedia.org/

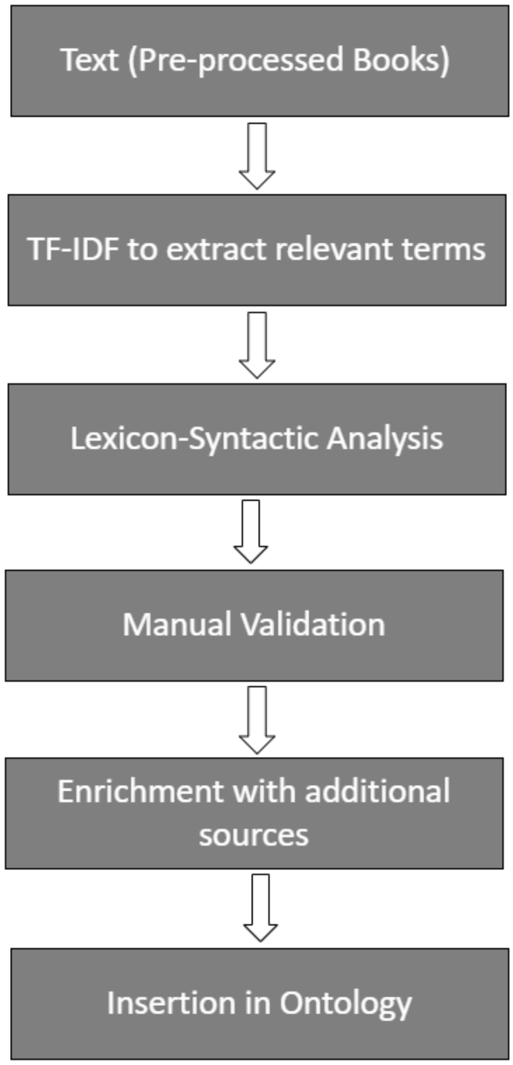

Figure 2.1: Our ontology refinement technique

TF-IDF for information extraction. This step extracts relevant domain terms from the bag of words generated from the pre-processing step. To this end, we applied the TF-IDF algorithm [26] to detect and extract the most relevant words in the context of the input content (textual books). The TF-IDF algorithm detects words that give more information to the domain. It is worth noting that, there are compound words that are equally relevant to the context. In order to obtain those words, we applied the algorithm not only to the unique words in the text, but also to Bi-grams and Tri-grams contained in the under analysis texts. This is justified due to the fact of not only extracting relevant words, but to capture terms associated with the context of the candidate word. At the end of this stage, we obtain candidate terms to be inserted in the ontology.

Syntactic analysis. With the candidate terms to be inserted in the ontology, this step aims to extract possible triples (subject, predicate and object) between the candidate terms and terms of the analyzed context. For this purpose, we apply the *POS-Tagging* [21] algorithm to classify expressions that contain the candidate terms from the previous step regarding the grammatical syntactic class. Then, the candidate terms with the respective classes were analyzed as in Arnold *et al.* [2], which relations are extracted according to the type of the semantic relation existing in an expression. For example, in the expression "Car is an automobile", the entity "Car" is the subject of the triple, "is an" is a pattern of Hyperonymy relationship and "Automobile" is an object. Therefore, for the candidate term "Car", the triple: Car- is one- Automobile is generated. At the end of this stage, a set of triples ready to be inserted in an ontology are generated with the candidate terms.

Validation. In this step, the generated candidate terms relevant to the domain are checked for errors. Terms that do not make sense to the domain, wrong terms are removed from the set. This process is manually performed by the authors and ensures the correctness of the operations.

Vocabulary enrichment with additional sources. The enrichment step takes into account additional Lexicon Database sources. In particular, this work explored the *BabelNet*[5]. This background knowledge is consulted to add synonym words for the candidate concepts and validates the relations extracted in the syntactic analysis. Other hyponyms obtained from the *BabelNet* are also inserted in the set.

Matching terms with the ontology. This step verifies where the candidate terms are inserted in the ontology. To this end, we designed and implemented an algorithm to find the relationship between existing entities in the ontology and the candidate terms. The algorithm works as the following: If a candidate exists in the ontology, its synonyms and/or hyponyms are inserted in the ontology; if the candidate do not exist in the ontology, the algorithm inserts its hyponym in the ontology as a broader entity, and simultaneously, the candidate and its synonyms are inserted as a specific entity; if both the candidate term and its hyponyms or synonyms do not exists in the ontology, the algorithm searches in the ontology the words caught in the context (cf. Step syntathic analysis) and inserts the candidate as "related to " the word in the context presented in the ontology.

Insertion in the Ontology. At this point candidates terms are ready to be inserted in the ontology as new triples. Important to note that, the ontologies considered in this

[5]babelnet.org/

work are in the SKOS [6] schema. This implies that, the relations in the ontology respects this schema and need to be adapted. In this step, we build the triples considering the underlying ontology schema and insert them into the ontology.

2.4 Metadata extraction on video lecture subtitles

We conducted experiments to assess the quality of semantic annotations obtained from a set of real-world video lectures available on *Youtube*. Our aim was to compare the effectiveness of existing tools in this task considering automatic transcription of audio videos. We investigated the quality of annotations in our study context with different ontologies to evaluate the impact of specific and general-domain ontologies. Moreover, we evaluate our results relying on our ontology refinement technique by comparing the results obtained from a refined ontology V_2 (via the use of our technique – cf. Section 3.3.2) and the input original ontology ($CSO\ V_2$).

2.4.1 Methodology

The analysis consisted of extracting the textual transcriptions from video lectures and performing the semantic annotation task. We considered a set of video lectures in Computer Science area. The semantic metadata extraction process automatically retrieved the subtitles from a video lecture in a textual format. The procedure used these textual subtitles as input for the semantic annotation tools.

The video lectures were from Computer Science considering the subjects of Computer Architecture, Data Structure and Computer Networks as scenarios in the experiments. Our experiments relied on both videos in Portuguese and English language with the use of the *Youtube API*[7] for automatic caption generation. The captions of videos in Portuguese Language were automatically translated for the purpose of only using textual transcriptions in English. At this moment, we aimed to evaluate the effectiveness of automatic textual transcriptions and translations from the video audios as input for the semantic annotation task. Table 2.1 shows the chosen videos.

For the development of our ontology refinement technique, we used the languages *Java* and *Python* to implement the algorithms. We used the *PyPDF2* library of *Python* for the extraction of texts in *PDF*. For computing *TF-IDF* and *POS-TAGGING* techniques, we used the *nltk* library of *Python*. Finally, we used the *API* from *BABELNet*[8] to apply the refinement using additional external data sources.

We stressed the use of distinct ontologies as a basis for the annotation tools. Our procedure considered three different ontologies. The first one was the *DBpedia* ontology[9]. It is a shallow, cross-domain ontology, which has been manually created based on the most commonly used infoboxes within Wikipedia. The second one was the Computer

[6]https://www.w3.org/TR/2008/WD-skos-reference-20080829/skos.html
[7]https://developers.google.com/youtube/v3/docs/captions
[8]`https://babelnet.org/`
[9]https://wiki.dbpedia.org/services-resources/ontology

Table 2.1: Video instances used to perform the experiments.

ID	Topic	Language	Duration	#Words	YouTube's ID
1	Computer Architecture	Portuguese	1:38:14	5858	2PtkVHCF1eE
2	Computer Architecture	Portuguese	0:35:21	5056	dQ6dzOPY9uc
3	Computer Architecture	English	0:13:53	2361	So9SR3qpWsM
4	Computer Architecture	English	0:12:16	2144	fvN98a7AT4
5	Computer Architecture	English	0:12:11	1952	UWDzhz8MVqc
6	Data Structures	Portuguese	0:40:16	6065	G0OKWQN9Jt4
7	Data Structures	Portuguese	0:33:34	4943	FzPceEhQCSQ
8	Data Structures	English	0:52:39	7129	9Jry5-82I68
9	Data Structures	English	0:52:31	6739	B7hVxCmfPtM
10	Computer Networks	Portuguese	0:38:40	4683	TWfqf-0gBgM
11	Computer Networks	Portuguese	0:37:08	4696	QjPTxELQvCg
12	Computer Networks	English	0:07:21	1095	1z0ULvgpW8
13	Computer Networks	English	0:07:06	1006	NX99ad2FUA

Network Ontology[10] available at *NBCO portal* which has about five hundreds classes. The third one was the Computer Science Ontology (CSO)[28]. We considered three versions of this Computer Science Ontology: one named as CSO V_1 presenting approximately 26 thousand entities; a refined version named CSO V_2 with 14 thousand entities; and our refined ontology from the CSO V_2 named as Refined Computer Science V_2^R with 15 thousand entities. These ontologies might enable us to understand the effects of different types and quality of ontologies as a support in the annotation process.

This study concerned some key main features for the evaluated semantic annotation tools: 1) API exposure to provide an available implementation of our process. In this sense, it is possible to (re)conduct as many tests as desired; 2) Flexibility on the used ontology. As the semantic annotation tools are only able to annotate terms which match the used ontology's concepts, if the tool supports more than one ontology, it is possible to assess different ontologies and determine which one is the best suited for annotating videos in a specific subject. We setup the annotation software tools as a preparation step to run the tests as follows.

AutoMeta. We selected four setups. The first one was running the tool with the *DBPedia* ontology; the second one was to run the tool with the Computer Science ontology V_1; the third one considered the Computer Science ontology V_2; and the fourth run the tool with the version V_2^R. In order to keep consistency to analyze the results, no reasoner was used for any evaluation tests.

CSO-Classifier. In our experimental test, we applied all versions under study of the Computer Science Ontology [28] (V_1, V_2 and V_2^R) to compare results among the different annotation tools.

NCBO. It has a series of setup parameters. For example, whether it should remove stop words; whether including ancestors of the annotated class; whether it should exclude numbers, *etc.*[11]. Our experimental evaluation with this tool used the default values for

[10]https://bioportal.bioontology.org/ontologies/CN

[11]http://data.bioontology.org/documentation

all the parameters except for the ontology. The set of ontologies accepted by the tool is restrict to the ontologies available at its tool's portal. We set the Computer Network ontology[12]. The decision of using this ontology was to assess another ontology related to Computer Science available at the tool's portal and evaluate the influence of different ontologies on the results.

OntoText. This tool uses only the *DBPedia* ontology as the basis to obtain the annotations.

A total of eight setups were considered to conduct the evaluation: (1) AutoMeta with DBPedia ontology; (2) AutoMeta with Computer Science ontology V_1; (3) AutoMeta with Computer Science Ontology V_2; (4) AutoMeta with the Refined Computer Science V_2^R (At this stage, we applied our proposed ontology refinement technique based as input the Computer Science V_2. We used computer science books as input content for the refinement); (5) CSO-Classifier with the Computer Science ontology V_1; (6) CSO-Classifier with the Computer Science ontology V_2; (7) NCBO with the Computer Network ontology; (8) Ontotext with DBPedia ontology.

For replication purpose, we turn available our developed software codes to thoroughly run this study[13].

Metrics. We defined metrics for result analysis. First, we obtained a set of terms that describe each video well. This set was built by running all tools for each video and get all distinct annotated words. Then, this set of all distinct annotated words was manually scanned. Those terms considered irrelevant was removed from the set. The remaining terms represent the set of relevant terms. The size of the set for each video is represented by DRV.

Precision (Pr) represents the precision of the tool on annotating relevant terms. It refers to the fraction of distinct relevant terms that were correctly annotated by the tool among all distinct annotated terms.

$$Pr = DRA/DTA \tag{2.1}$$

DRA = number of distinct (without repetition) relevant terms that the tool annotated **correctly** (according to the set of relevant terms).

DTA = total number of distinct (without repetition) terms that the tool annotated (it includes the terms that are not relevant according to the set of relevant terms).

DRV = total number of distinct relevant terms for a given video.

Recall (Re) represents the tool's ability to recognize the relevant terms for that video. It refers to the fraction of distinct relevant terms that were correctly annotated among all distinct relevant terms for that video.

$$Re = DRA/DRV \tag{2.2}$$

We used F-Score (F) as the harmonic mean of precision (Pr) and recall (Re) to rank the tools in our result analysis.

[12]https://bioportal.bioontology.org/ontologies/CN

[13]https://gitlab.ic.unicamp.br/jreis/semantic-metadata-extraction

2.4.2 Experimental Results

We organized the results by each subject presented in the videos. In this sense, all results concerned with Computer Architecture, for instance, are analyzed together. Tables 2.2, 2.3 and 2.4 present the achieved results by the studied setups for each subject.

Table 2.2 shows that *AutoMeta* with the *DBPedia* ontology got the highest f-score and recall (0.56 and 0.80). Results show that the precision for the tools using the Computer Science ontology V_2 was considerably high (0.30 for *CSO-Classifier* and 0.41 for *AutoMeta* with V_2; and 0.65 concerning the refined ontology V_2^R). However, the recall value for these tools was lower, which negatively affected the f-score. Considering the number of distinct relevant terms annotated, *AutoMeta* with *DBPedia* ontology had the best effectiveness by annotating 121 terms.

Table 2.3 shows that the use of our Refined Computer Science V_2^R ontology got the highest number in f-score (0.53 with *Autometa*). This shows that the precision for those tools using CSO V_2 was high (0.70 for *CSO-Classifier* and 0.71 for *AutoMeta*). However, the recall for these tools influenced the f-score in a negative way. Both *Ontotext and NCBO* had the lowest recall (0.19 and 0.06, respectively) and lowest f-score (0.21 and 0.13, respectively). Considering the number of distinct relevant terms annotated, *AutoMeta* with *DBPedia* ontology had the best effectiveness by annotating 121 terms. *AutoMeta* with CSO V_1 was the second best because it was able to annotate 99 relevant terms.

Table 2.4 reveals that the *CSO-Classifier* with CSO V_2 got the highest f-score (0.56). This result presents that the precision for *NCBO* tool was the highest (0.83). In the computer networks subject, the tools with *DBPedia* got the lowest results. This is verified by the number of distinct relevant terms correctly annotated (DRA)(17 for *AutoMeta* with *DBPedia* and 35 for *Ontotext*).

We present the overall results by each setup with average results and confidence interval for the mean of 95%, representing an interval of plausible values for population mean. Table 2.5 presents the obtained results to analyze the overall effectiveness for each setup. We calculated the overall precision, recall and f-score for all videos under analysis. The results shows that CSO-Classifier got the highest f-score in an overall (0.46). However, the recall for all the tools were low which decreased their overall f-score. Considering the total number of relevant terms correctly annotated (DRA), the setups that used *DBPedia* and CSO V_1 ontologies had the best effectiveness by correctly annotating 62 and 50 terms, respectively.

2.5 Discussion

This study assessed a semantic metadata extraction process based on automatic caption generation and ontology-based semantic annotation tools. This investigation aimed to analyze the effectiveness of key approaches in the entity recognition task in the context of video lectures annotation. Our evaluation explored a dataset of video lectures from Computer Science themes. We examined the influence of distinct ontologies as the basis for the annotation tools to analyze their effectiveness in the studied scenarios. Finally, this study compared the effects of our ontology refinement technique in the annotation

Table 2.2: Results for Data Structure subject videos. This table presents the tool's name, ontology used, distinct relevant terms for the videos (DRV), distinct terms annotated (DTA), distinct relevant terms correctly annotated (DRA), precision (Pr), recall (Re) and f-score

Tool	**Ontology**	**DRV**	**DTA**	**DRA**	**Pr**	**Re**	**F-Score**
AutoMeta	DBPedia	178	281	121	0.431	0.801	0.560
AutoMeta	Computer Science V_1	178	171	32	0.187	0.212	0.198
AutoMeta	Computer Science V_2	178	51	21	0.412	0.139	0.207
AutoMeta	Refined Computer Science V_2^R	178	60	25	0.65	0.245	0.37
CSO-Classifier	Computer Science V_1	178	82	22	0.268	0.14	0.183
CSO-Classifier	Computer Science V_2	178	49	15	0.306	0.096	0.146
Ontotext	DBPedia	178	70	15	0.214	0.099	0.135
NCBO	Computer Network	178	22	3	0.136	0.02	0.034

Table 2.3: Results for Computer Architecture subject videos. Table presents the tool's name, ontology used, distinct relevant terms for the videos (DRV), distinct terms annotated (DTA), distinct relevant terms correctly annotated (DRA), precision (Pr), recall (Re) and f-score

Tool	**Ontology**	**DRV**	**DTA**	**DRA**	**Pr**	**Re**	**F-Score**
AutoMeta	DBPedia	278	294	122	0.415	0.457	0.434
AutoMeta	Computer Science V_1	278	267	99	0.371	0.371	0.371
AutoMeta	Computer Science V_2	278	121	87	0.719	0.326	0.448
AutoMeta	Refined Computer Science V_2^R	278	140	97	0.358	0.47	0.53
CSO-Classifier	Computer Science V_1	278	198	72	0.364	0.266	0.307
CSO-Classifier	Computer Science V_2	278	134	94	0.701	0.347	0.464
Ontotext	DBPedia	278	170	51	0.300	0.191	0.233
NCBO	Computer Network	278	32	16	0.500	0.060	0.107

Table 2.4: Results for Computer Network subject videos. Table presents the tool's name, ontology used, distinct relevant terms for the videos (DRV), distinct terms annotated (DTA), distinct relevant terms correctly annotated (DRA), precision (Pr), recall (Re) and f-score

Tool	**Ontology**	**DRV**	**DTA**	**DRA**	**Pr**	**Re**	**F-Score**
AutoMeta	DBPedia	220	242	17	0.070	0.082	0.075
AutoMeta	Computer Science V_1	220	214	50	0.234	0.242	0.237
AutoMeta	Computer Science V_2	220	103	44	0.427	0.213	0.284
AutoMeta	Refined Computer Science V_2^R	220	140	47	0.456	0.43	0.32
CSO-Classifier	Computer Science V_1	220	190	86	0.453	0.41	0.430
CSO-Classifier	Computer Science V_2	220	144	100	0.694	0.476	0.564
Ontotext	DBPedia	220	181	35	0.193	0.169	0.180
NCBO	Computer Network	220	80	67	0.838	0.324	0.467

Table 2.5: Overall results for the tested setups. Table presents the tool's name, ontology used, average distinct relevant terms with 95% confidence interval (DRV), average distinct terms annotated with 95% confidence interval (DTA), average distinct relevant terms correctly annotated with 95% confidence interval (DRA), precision (Pr), recall (Re) and f-score

Tool	**Ontology**	**DRV**	**DTA**	**DRA**	**Pr**	**Re**	**F-Score**
AutoMeta	DBPedia	49 [36; 61]	62 [44; 81]	20 [18; 30]	0.318	0.416	0.360
AutoMeta	Computer Science V_1	49 [36; 61]	50 [26; 64]	13 [7; 20]	0.278	0.290	0.283
AutoMeta	Computer Science V_2	49 [36; 61]	21 [13; 28]	11 [5; 17]	0.553	0.243	0.337
AutoMeta	Refined Computer Science V_2^R	49 [36; 61]	60 [26; 62]	15 [7; 25]	0.35	0.285	0.40
CSO-Classifier	Computer Science V_1	49 [36; 61]	36 [22; 49]	13 [8; 18]	0.383	0.282	0.324
CSO-Classifier	Computer Science V_2	49 [36; 61]	25 [14; 36]	15 [7; 23]	0.633	0.324	0.428
Ontotext	DBPedia	49 [36; 61]	32 [20; 44]	7 [4; 10]	0,193	0.169	0.180
NCBO	Computer Network	49 [36; 61]	10 [5; 15]	6 [2; 10]	0.838	0.324	0.467

process.

Results indicated that the ontology used by the software tools plays an important role in the task of annotating terms in a video lesson. Greater coverage of concepts that match relevant terms in videos leads to better results. This implies that specific ontologies of Computer Science would lead to better results than general domain ontologies such as *DBPedia*, due to the subjects covered in the considered videos.

The effects of an ontology refinement were analyzed in the experiments by comparing the results obtained with *AutoMeta* tool and *CSO-Classifier*, using both versions V_1 and V_2 of the Computer Science Ontology to annotate the videos. In these setups, we noted that the number of distinct terms annotated (DTA) and the number of distinct relevant terms (DRA) decreased in general. However, the overall results for precision and recall were higher using *CSO-Classifier* with Computer Science Ontology V_2. This effect is explained by the refinement of concepts represented in the ontology. In the tools using CSO V_1, more Computer Science related concepts are recognized, but, these concepts are not specific to describe the subject nor relevant to the video content.

Regarding the software tools, our findings indicated that *AutoMeta* obtained more consistent results. Their precision and recall were good in an overall as well as its ability to annotate several relevant terms based on the input text. *CSO-Classifier* using CSO V_2 achieved the best overall results, mostly because this tool has the ability to annotate composite concepts and correctly associate with the ontology entities. For example, the tool recognizes concepts like "local area networks".

Ontotext obtained the worst results even using the *DBpedia* ontology. However, this tool recognized several relevant terms. Our findings indicate that this tool is adequate for recognizing terms from the input text, but the number of distinct relevant terms it can recognize is low.

NCBO presented the worst number of concepts annotated among all. Clearly, the used ontology had a huge impact on its results producing a high overall precision (0.83), but its overall recall remained low (0.32). In Computer Network related videos, the tool recognized almost all key concepts presented in the content of the videos. That means, for this setup, the set of relevant words recognized (DRA) are very relevant, but not broad enough to compare with the set of total relevant words per video (DRV).

Another point is the quality of automatic transcriptions associated with the language spoken in the videos. We observed that the use of automatically generated captions did not heavily interfere with the quality of annotations. This procedure generated meaningful texts and recognized specific words of Computer Science field. By considering the annotations for Portuguese videos, the overall results were not negatively impacted by using automatically translated subtitles. Nevertheless, to the best of our knowledge, there is still very little research in the area considering the use of Portuguese or other languages in semantic annotations. This implies the need for translating the transcripts of the video before the process of semantic annotation.

Regarding the results obtained from our ontology refinement technique, we observed an increase of approximately 10% on average in the number of terms relevant to the set of videos before refinement (CSO V_2), and after refinement (*Refined Computer Science* V_2^R). This change can be observed when considering the number of different relevant terms within the total set of relevant terms. There was an increase in this metric considering the new refined version of the ontology applying our procedureThese results were expected because the refinement procedure improved the number of unique entities by 1 thousand (approximately 10%).

As impact for the evaluation in our study context (video lectures), the process of semantic annotation still suffers with the need of gold standards (baseline datasets) for comparison among different annotation tools. This leads to costly work on qualitative analysis of the semantic annotations obtained and the use of human annotators to validate the relevant terms recognized.

The key challenges for the future work on extracting metadata from video lectures are related to defining an ontology suited to hold the most relevant terms for the videos. The results demonstrated that using broad ontologies such as the *DBpedia* and the Computer Science ontology can give good results on extracting some relevant terms from the videos.

In general, our results showed that the *DBPedia* ontology contains useful concepts to describe the videos compared to the CSO ontology. We emphasize that the CSO ontology described broad themes in Computer Science, and does not delve into more specific concepts. The existence of ontologies for specific subjects, to which the video contents refer (Computer Architecture, Data Structures and Computer Networks) could lead to better annotation results. Our ontology refinement technique was able to improve the CSO ontology (CSO – V_2^R) to catch more of these specific concepts and was able to improve the results obtained from the previous version of the ontology (CSO – V_2). However, the need for manual validation step and good input from the book resources are still a limitation in this process. The extraction of natural language resources and the process for creating ontologies tends to consider language aspects such as ambiguous terms.

We consider the process for automatic extraction of relations from raw text is still a challenging task. On the basis of the results obtained from our experiments, we highlight that there was a direct improvement in the amount of relevant terms detected and the combination of ontologies for the semantic annotation process is still the ideal setup.

2.6 Conclusion

The extraction of semantic metadata in video lectures can be a key part to improve the recommendation and retrieval of videos in the learning process support. The annotated words can describe better the video's subjects and enable semantic correlation from one video to other. In this investigation, we described a technique to semi-automatically refine an ontology considering domain specific book sources. We applied this technique in the refinement of an existent Computer Science ontology. This study assessed the effectiveness of this method and evaluated it in the context of annotations obtained from real-world video lectures. Our conducted experiments with different semantic annotation tools and different ontologies point out that obtained annotations considering an ontology related to a specific domain can achieve more precise results. We found that less domain-specific ontologies like *DBpedia* can help in the process. The use of cross-domain ontologies should not be totally unconsidered in the process. Our experimental results were relevant to understand which parts of the whole metadata extraction process can influence the most the quality of the extracted metadata. Also, we found how domain specific concepts leads to better results in the metadata extraction process. Future work involves the improvement of our defined refinement process. In experimental aspects, we plan to add further book resources as the input for the process and reapply the refinement on the previous refined ontology.

Chapter 3

Video Lecture Recommendations based on Semantic Annotations

3.1 Introduction

The growth of information dissemination in Web has changed the availability of multimedia content for helping in the learning process. This results in a significant increase in the amount of educational resources accessible to learners [35]. In this context, learners need to collect the appropriate content that helps them in their learning process. There are contexts in which learners need to find similar online video lectures covering similar content. Potentially, video lectures from other courses or teachers can be interesting to replace or complement concepts under study.

The key challenge in this process is to find similar videos conceptually related with a topic (concept) in a given video lecture. In this context, recommendation systems serve as a solution to the problem of filtering most appropriate educational content in a personalized way. Recommendation systems are especially useful for users in online systems. In these systems, users encounter a big amount of information they need to handle manually. Recommendation systems play a key role in contexts in which users need to make decisions over a large set of options [6].

In this work, we conceptualize, develop and evaluate an automatic video lecture recommendation method that explores semantic annotations created from video lecture subtitles. We assume that the use of ontology-based annotations enables new ways of recommending educational content from video lectures. Our solution computes clusters among video lectures and combines it with semantic similarity computation for ranking video lectures given the request of a user. We provide a software tool working as a video lecture player. In this tool, learners can sign a doubt when watching a video and the system recovers semantically related online lectures based on the semantic annotations encoding the videos.

We conduct an experimental evaluation relying on forty real-world video lectures from the Computer Science domain and we observed how the combination of different techniques and the use of semantic annotations can improve the quality of the recommendations.

The remaining of this work is organized as follows: Section 3.2 describes fundamental concepts of this investigation and discusses related work; Section 3.3 reports on our proposed technique for video lecture recommendation; Section 3.4 describes the experimental results whereas Section 3.5 discusses our findings; Section 3.6 provides the conclusion remarks.

3.2 Background

We define relevant key concepts for this investigation followed by an exploratory analysis of related work.

3.2.1 Fundamental Concepts

Gruber [13] defined an ontology in the context of Computer Science as a formal, explicit and shared specification of a conceptualization. Conceptualization is understood as an abstract model; explicit implies that the elements must be clearly and objectively defined; and formal indicates that the specification must be interpretable by machines. This specification is theoretical and therefore requires a specific representation that allows the description of an entity without ambiguity for the detailed representation of domain concepts. The main advantage and applicability of ontologies refers to the representation of knowledge and the possibility of information reuse generated by other applications.

Recommendation systems aim to select personalized items to satisfy users' interests and needs [6]. According to Brunialti *et al.* [5], recommendation systems can be classified as how to filter information for content recommendation. In collaborative filtering recommendation systems, the recommendation is based on the history of items that users have interacted with it in the past. In recommendation systems filtered by content, recommendations are generated based on the characteristics of items. In knowledge-based recommendation systems, item suggestions are offered based on inferences about users' needs and preferences. There are also hybrid systems that combine different types to overcome the disadvantages of each category.

Recommendation systems with content filtering analyze the characteristics of candidate items by extracting attributes that describe these items. Knowing characteristics referring to items that a user consumed, a system can recommend new items to users [5]. In general, the task of recommendation systems filtered by content is to extract relevant characteristics from items acquired by a user and, based on this, compute similarity between all items in the system. On this basis, recommend the most similar items in relation to the user's preference.

Recent studies on recommendation systems explored automated approaches that involve the application of information retrieval techniques or machine learning [5]. There are limitations in the use of these methods regarding the way the system acts based on the lack of initial data on user's profile, or characteristics of new items to be recommended. In this sense, one of the major limitations is in the analysis of poorly structured data, in addition to the complexity of extracting and analyzing non-text content. Another difficulty related to this approach is to analyze the semantics in different contexts [1].

Techniques commonly used in recommendation systems are text mining techniques such as TF-IDF (Term Frequency - Inverse Document Frequency) [29] and Cosine Similarity [22]. These techniques consist in the computation of a numeric metric that classifies the similarity in a set of documents. On this basis, it is possible to determine a ranking among the most similar documents. TF-IDF technique [29] explores the concept of frequency of terms in a set of documents and the relevance of these terms within the set. These two metrics enable detecting the similarity between documents through vector distance between them.

Cosine similarity [22] is a technique that uses the counting of terms in common between documents and calculates the cosine of the angle between the two vectors projected in a multi-dimensional space. In this sense, the two vectors are representations counting the words between the documents.

3.2.2 Related Work

Recommendation systems exploring ontologies and/or taxonomies as a knowledge base have been developed in literature [10, 36, 16, 20, 38]. In this approach, information representation is enriched through the description of concepts presented in ontologies.

The most common approach in literature has developed personalized recommendation systems based on collaborative filtering [36, 16]. In this approach, information about users and the characteristics of recommended items are described using semantic information from ontologies modeled for the application domain. After the classification of items and users, similarity is calculated between users and their preferences. Afterwards, based on the correlation metric between item-user pairs, the systems provide appropriate recommendations.

Another approach takes advantage of the use of ontologies is called adaptive context-based recommendation systems [25, 20, 38]. This type of system analyzes and considers other factors such as: user experience, time or location, to outline a context and adapt recommendations for users. In this type of application, ontologies are used for determining the appropriate context for the recommendation. Ontologies are relevant to specify not only related concepts, but they represent contextual information in recommendations, such as the time that a particular item was recommended or the usefulness of the recommendation to the user. For example, in the system proposed by Yu *et al.* [38], applied to the virtual learning area, the model helps in the process of providing resources for students. This system generates context-based recommendations relying on the use of ontologies to represent knowledge about users, the contents and the application domain. In this work, recommendation results are different depending on factors such as the student's level of experience and progress in the course at the time the recommendation was made.

Our research differs from existing work in literature due to the way ontologies are used to describe concepts and how the recommendation is produced on this basis. In our approach, we developed and applied a refinement technique in the ontology based on content extracted from book resources related to the system application domain. At this point, we analyzed the effects of this ontology on the classification of contents, regarding the specificity of the concepts obtained and the precision of the recommendations provided. In our recommendation method (cf. Section 3.3), we consider ontology to recover related concepts. Our method applies not only the knowledge obtained from the ontology, but it provides a combination of recommendation techniques to enrich our recommendation mechanism. When a student assign a doubt to the system when watching a video, our system detects ontology concepts related to the content being communicated at that time of the video lecture. Our method explores generated ontology-based semantic annotations for this purpose.

3.3 Recommendation Method

We present the general architecture designed for our semantic-enhanced video lecture recommendation system (cf. Subsection 3.3.1) followed by the details of our implemented recommendation technique (*cf.* Subsection 3.3.2).

3.3.1 Architecture

We designed and developed a method to recommend video lectures in a given subject, considering related topics in a online video class being watched by a learner. The system allows the student to indicate an option of doubt at some point in a video. At this stage, our system recommend specific excerpts of other videos containing related concepts to the subject of doubt in that specific moment. For this purpose, ontology concepts from semantic annotations are explored for the indexing of videos, as well as content recommendation techniques perform a semantic retrieval task of related videos.

Our proposal is organized into three several steps: (1) transcription of video content; (2) semantic annotation of video content subtitles based on available ontologies;and (3) recommendation of video lectures based on the concept found in the segment of the video signaled by a user. Figure 3.1 presents our defined method to obtain semantically annotated videos and the process for recommending related videos to the user.

Transcription of video contents. With the aim of semantically classifying videos and indexing them, the first step generates transcriptions from input online videos. This step considers a set of video lectures as input (A in Figure 3.1) and explores automatic generation of captions to extract textual transcription from videos (B in Figure 3.1). For this purpose, videos lectures are obtained from *Youtube* via the use of the *Youtube API* in order to automatically generate transcription from videos. Textual description considers the video subtitles as natural language texts.

Semantic annotation. Considering the textual transcription obtained from videos (C in Figure 3.1), semantic annotation techniques (D in Figure 3.1) are applied to achieve semantic metadata from video contents (subtitles). The semantic annotation process recognizes concepts in the input transcription from existing ontologies (E in Figure 3.1). At this stage, we explore the use of existing semi-automatic tools to perform this task. These tools assist the extraction of semantic annotations from textual resources based on ontologies. Experiments conducted in Borges *et al.*[4] report on results of applying existing semantic annotation tools to real data and the challenges in addressing the automatic

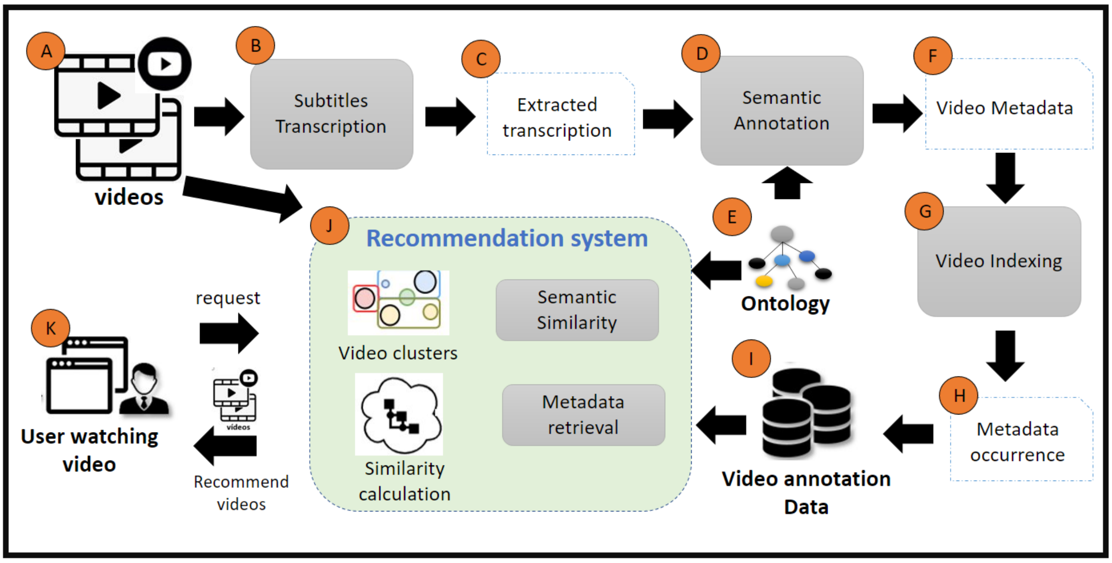

Figure 3.1: Method for semantic metadata extraction and recommendation of video lectures.

creation of semantic metadata from video subtitles. At the end of the annotation process, the system recognizes and extracts entities (F in Figure 3.1) in a video transcription and connects them to their respective concepts presented in the available ontology. This produces the video metadata (F in Figure 3.1).

Video indexing. This step refers to the video indexing (G in Figure 3.1). In this process, the annotated entities from videos are associated with their respective timestamp of occurrence in the videos. To this end, terms annotated in the videos and the respective subtitles are compared. In this process, each entity occurrence in an exact point at time of the video (in the subtitle) is identified and associated to the produced annotation. The output of this process is the metadata occurrences embedded in the subtitles (H in Figure 3.1). This contains information about each entity in addition to the time it occurs in videos. This data is stored in a database representing data semantically associated to video lectures (I in Figure 3.1). This database is useful to manage the retrieval of specific contents in a given video to support our recommendation mechanism.

3.3.2 Recommendation Mechanism

We designed a recommendation mechanism based on the created database of video metadata and a user interaction interface for video watching and signing of doubts. The functioning of the system works as follows: first, a user makes a request to the system informing a doubt in a part of a video during his/her interaction with the system while watching a video. Then, the system is responsible for recovering the topic (concept annotated in the subtitle) in that specific part of the video. On the basis of the detected concept, the system returns an ordered list of video lectures related to the identified concept of doubt, retrieved from the database. As the final result, the user has access to a list of videos conceptually aligned to the topic of doubt.

In the recommendation technique, we consider a content filtering approach in which

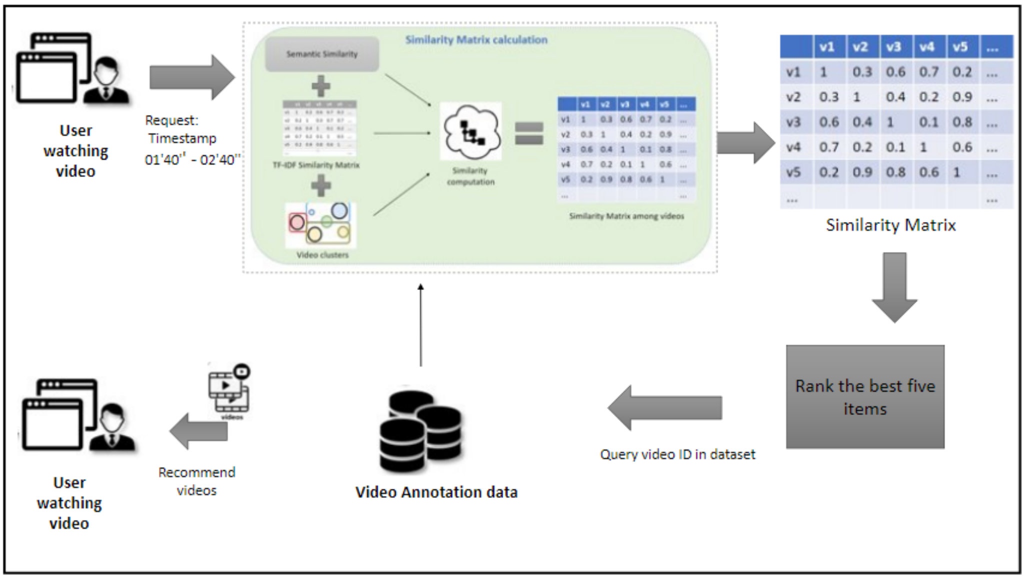

Figure 3.2: Recommendation Engine.

videos are recommended based on the similarity computed between the annotated metadata from videos. Ontology is used in this context to relate concepts associated with the topic of doubt, taking advantage of the structure and semantics expressed in ontology. We consider a combination of information retrieval techniques, such as TF-IDF and cosine similarity, unsupervised machine learning and semantic similarity of annotated concepts to achieve improved content recommendation (*cf.* Figure 3.2 describes the technique).

The recommendation technique requires as input the following elements: 1) set of concepts semantically annotated in a certain time of video; 2) the set of videos that these concepts appear in the system database.

We considered three aspects in our recommendation technique: 1) The diversity of recommendations; 2) the similarity between semantic concepts annotated considering the ontology; and 3) The clustering between videos. We defined a metric that combines these three different aspects by exploring the following techniques: 1) machine learning for clustering related videos; 2) TF-IDF as an information retrieval technique for extracting textual content; and 3) computing of similarity between concepts expressed in the ontology.

The machine learning technique applied was the *K-Means* algorithm proposed by Harmerly and Elkan [14]. In particular, the entire set of videos is grouped based on k[1] main classes. We recommend the number of classes related to the number of key topics from the video database. In particular, our study context with video lecture in the Computer Science area (cf. Evaluation results in Subsection 3.4), we considered four classes, such as: "Data structure", "Computer networks", "Computer architecture" and "Programming Language". Therefore, initially the *K-Means* is executed with the entire set of video annotations to associate them into one of these clusters. In our approach, if two sets of videos have the same group, they are generally correlated.

At this stage, our technique aims to detect to which extent a set of concepts from

[1] In our experiments, we used $K = 4$.

a given moment of a video (based on the user's point of doubt) is related to the most similar videos in the database. For this purpose, TF-IDF is applied to the set of annotated concepts in videos database to compare the similarity between a set of concepts annotated in a given time of the video lecture signaled by a learner, and the videos in the database. We note that this step is performed each time a user request a recommendation to the system. In this sense, the TF-IDF similarity matrix is generated between a set of concepts in a given time and the entire set of videos. At the end of this process, the TF-IDF similarity matrix is used in the equation to generate our recommendation matrix

We explore a metric to compute the semantic similarity between the most frequent concepts between two sets of annotated concepts in videos. We use the Ontology-Based Semantic Similarity equation defined in [37, 12] to calculate the similarity between concepts expressed in an ontology. The techniques works based on the distance of these concepts.

$$OS(c1, c2) = 2H/(D1 + D2 + 2H) \tag{3.1}$$

Where $c1$ and $c2$ are the most frequent concepts between two sets of annotated terms; $D1$ and $D2$ are, respectively, the shortest path from the term $c1$ and $c2$ to the deepest common node in the ontology; and H is the shortest path from the deepest parent node in common to the root concept of the ontology.

This metric is used in our system to compare the most frequent concept in a set of concepts from a given moment and the most frequent concept for each video. On this basis, we assume that the most frequent concept in a semantic annotation represents the key content being approach in that moment of the video and we calculate how much those concepts are related to each other based on the ontology.

Our video lecture recommendation system considers the metric that defines the similarity between the terms annotated at a given moment in a video lecture and each video from the database as the result of the following equation:

$$Sim(Ta, Vi) = 0.1 * CL + 0.5 * TF - IDF + 0.4 * OS \tag{3.2}$$

Where Ta represents the set of terms annotated at a given moment in a video; Vi represents each video lectures from the videos database;

- CL signals if the pair Ta and Vi are in the same cluster (*cluster*): 0 if they are not in the same cluster and 1 if they are in the same cluster.
- $TF - IDF$ represents similarity value based on the TF-IDF method;
- OS represents the semantic similarity between the most frequent terms in Ta and Vi in relation to the ontology.

Our proposed equation is used in the recommendation to fill a similarity matrix between the set of annotated concepts of a given moment of doubt and the entire set of video lectures in the database. In other words, each time a request is made by the user, the system generates a similarity matrix between the set of concepts and the other videos.

Important to mention that our similarity equation metric considers weights for the three techniques applied as a way to assign different importance to each of the techniques considered in the equation. This combination of weights was considered to have the best impact in penalizing the diversity of recommendations whereas considering the similarity of related videos, and considering the semantic aspects of our recommendation method.

For each online video lecture in our system, our method produces a similarity matrix representing the similarity value between the annotated terms of that video and the video lectures in the system. Based on the generated matrix, a doubt request from a learner at the moment t of a video, the system returns an ordered list of most similar videos from the database of video lectures by getting the best values for the similarity measure.

3.4 Evaluation

This section presents how our system was implemented (cf. Subsection 3.4.1). Subsection 3.4.2 presents the conducted procedure and dataset in our evaluation followed by the obtained results (cf. Subsection 3.4.3).

3.4.1 Implementation and System Interface Prototype

For the development of our proposed method, we explored existing software tools to provide an available implementation. We investigated the use of the *API* of *Youtube*[2] for automatic caption generation. This *API* was used in conjunction with existing tools for semantic annotation such as *Autometa* [3], *Ontotext* [4] [17] and NCBO Annotator[23]. Our implemented codes developed in this research are available in the public repository [5] [4].

For the development of the recommendation system, we used the languages *Java* and *Python* to implement the algorithms. The development of the video lecture player system, the *Javascript* language was used in conjunction with the *API* of *Youtube* [6] that allows us to embed a *player* video from *YouTube* on a Website. We used the *Json* format to store the semantic annotations associated with video timestamps stored in a *MongoDB* database.

System Interface. To make the system interactive and easy to use by learners, we developed an user interface to run the system which plays a video from the *Youtube* and presents a button where a learner can request a recommendation to our system. The interface is able to shows the *Youtube* links to the recommended videos, where the user can input the Link into the interface and play the video or access the link in the Web Browser (*cf.* Figure 3.3).

The interface work as follows: Initially the user Input the *Youtube* Video Link into the space (*cf.* element 4 in Figure 3.3) and load that into the video player (*cf.* element 1 in Figure 3.3). If the video do not exist in our database or was removed from the

[2] `https://developers.google.com/youtube/v3/docs/captions`
[3] `https://github.com/celsowm/AutoMeta`
[4] `https://ontotext.com/business-cases/content-packaging-reuse/`
[5] https://gitlab.ic.unicamp.br/jreis/semantic-metadata-extraction
[6] `https://developers.google.com/youtube/iframe_api_reference`

Figure 3.3: Recommendation System interface

Youtube, the system shows an error message and does not reproduce the video. If the user has a doubt and want to get some recommendation to help him/her when watching a video, (s)he can press the button (*cf.* element 2 in Figure 3.3) which makes a request to the system. At this stage, the system triggers the recommendation engine. After a few seconds, the interface shows the recommended video links (*cf.* element 3 in Figure 3.3).

3.4.2 Datasets and Procedure

We conducted an evaluation to assess the quality of our recommendation system results. We considered a set of real-world video lectures and evaluation metrics to assess its effectiveness.

We used a set of forty video lectures in our database which cover four sub-areas of Computer Science field: "Data Structure", "Computer Architecture", "Computer Network" and "Programming Language". All recovered videos were chosen randomly by the authors of this work considering only these main sub-areas. We considered video lectures in Portuguese language and the length of the videos were between 5-10 minutes of duration. On this basis, we ensure that different aspects of recommendation can be evaluated such as: the diversity of recommendation and the relevance of the content.

Our experimental procedure defined an ideal set of recommendations for 20 videos in the database, considering timestamps of 60 seconds length for each video. It generates around 100 recommendations lists to be compared with the results obtained by our results. The ideal set of recommendations were generated by one of the authors in this work considering the following main aspects: How the videos are relevant to the theme approached at the moment of the video; if the concept being approached is related to a sub-area of the video lecture; and if the recommendation would be redundant to be recommended. Based on these aspects, we created a list of recommendation ranked by

$$\text{diversity}(i_1, \ldots, i_n) = \frac{\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} (1 - \text{sim}(i_i, i_j))}{\frac{k(k-1)}{2}}.$$

Figure 3.4: Diversity Metric [18].

similarity and relevancy to the concept. Based on this list, our experiments measure the effectiveness of recommendation technique.

In our recommendation task, we define the following variables for computing *Precision* and *Recall* as follows:

- **REL-REC**: Number of video lecture recommendations generated by our mechanisms and judged relevant by evaluators (authors).
- **LIST-REC**: Number of all video lectures recommended by our mechanism in a given user request.
- **LIST-REL**: Number of all relevant video lectures expected by a user request.

On this basis, We compute *Precision* and *Recall* of a request as follows:

$$Precision = REL - REC/LIST - REC \tag{3.3}$$

$$Recall = REL - REC/LIST - REL \tag{3.4}$$

Precision concerns how many recommendations are relevant among the provided recommendations. Recall concerns about how many recommendations are provided among all the relevant recommendations expected.

In our experimental procedure, we considered a set of five recommendation for each request of doubt. On this basis, we computed three main metrics: MAP@K (Mean Average **Precision** over k different recommendations); MAR@K (Mean Average **Recall** over k different recommendations); and Diversity[18].

MAP@K and MAR@K were calculated as an average (mean) of average *Precision* and *Recall* of recommendations for the best k recommendations. For instance, for 100 video lectures, MAP@K is the sum of average Precision for all recommendations divided it by 100.

An ideal system should display some degree of diversity in the presented recommended items. We evaluated our system by observing the **diversity** metric. We measured how much video lectures in a recommendation list are similar. If the recommendations are too similar, they can be redundant to the user and not be of interest. Diversity can be considered as the opposite of similarity. Smyth and McClave [18, 7] defined diversity in a set of items, $c1..cn$, as the average dissimilarity between all pairs of items in the itemset. They introduced the following formula for measuring diversity:

To evaluate the diversity score of our recommendation system, we considered two different scenarios of experiments: First scenario recommending entire video lectures for each recommendation; the second scenario recommending specific parts of video lectures which are related to the concept being approached (at the time of doubt assigned by the user). To measure the second scenario we considered the same criteria as the first scenario, so we compared the specific parts recommended as being part of the ideal set of recommendation generated for the experiments. If the recommended part of the video is part of the ideal video recommendation, so that part is relevant too.

3.4.3 Experimental Results

We organized the results by each sub-areas of Computer Science under study presented in the videos. In this sense, all results concerned with “Programming Language”, for instance, are analyzed together. Table 3.1 presents the achieved results for the different sub-areas.

Table 3.1: Recommendation results for Mean Average Precision and Mean Average recall over different sub-areas videos.

Sub-area	**MAP@5**	**MAR@5**
Data Structure	0.764	0.336
Computer Architecture	0.723	0.241
Computer Network	0.649	0.458
Programming Language	0.843	0.540

Results indicate that the recommendation over the different sub-areas behaved similar for all these sub-areas. “Programming Language” sub-area obtained the best results for best 5 recommendations on average (0.843 for MAP@5 and 0.540 for MAR@5). This result means that our technique got more than 4/5 of relevant recommendations on average. It got around 2/4 of correct recommendations among the relevant recommendations expected. It shows that our system can get the similarity of videos, but can has space for improvements to identify all relevant ones.

Table 3.2 shows results regarding the different scenarios. It shows that Scenario 1 – where the entire videos are recommended got better results. In general, the system hits almost 80% of recommendations correctly by comparing to the ideal set of recommendations; and 37% of the recommended videos were provided among all the relevant expected ones. Other interesting effect observed is that the Scenario 2 – where only parts of the video were recommended got a better value for the Mean Average Recall (almost 50% of provided recommendations are indeed relevant to the concept being approached).

Table 3.2: Recommendation results for Mean Average Precision and Mean Average Recall compared between different scenarios.

Scenario	**MAP@5**	**MAR@5**
Scenario 01 - Entire Video recommendations	0.778	0.375
Scenario 02 - Timestamps video recommendations	0.571	0.469

Figure 3.4 compares the diversity score for the two scenarios analyzed. Results indicate the effects of recommending content considering a specific part of a video. Scenario

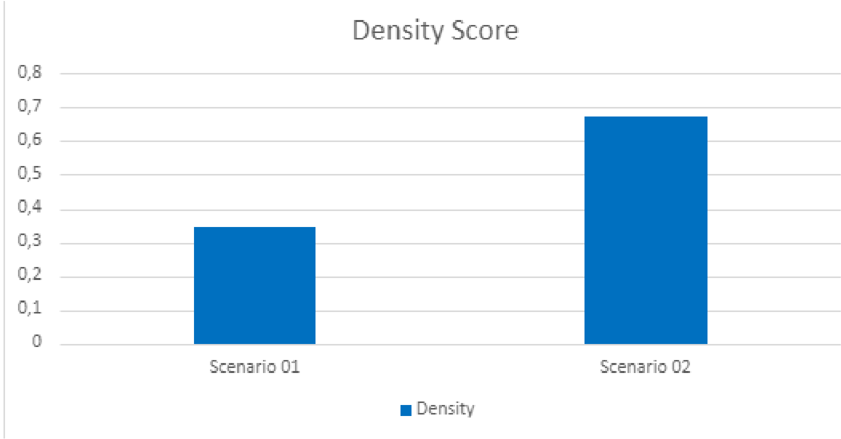

Figure 3.5: Diversity Score for both scenarios

2 obtained a score of around 0.7 for the diversity indicating that our system could recommend different videos from the database in such a way that not only videos directly related to the concept are recommended, but videos that in some way are from other sub-area.

3.5 Discussion

This research proposed a video lecture recommendation technique based on semantic metadata extraction process relying on a combination of Machine Learning, Information Retrieval and Semantic Web techniques. In our developed system, we aimed to achieve a technique to find similar videos related with a topic being approached in a video lecture at a given time t. Addressing it reveals several issues because language aspects in natural language tasks and the process of recommending relevant content to users based on different parts of videos are challenging.

One of the key relevant contributions of this research was the design of our video lecture recommendation technique taking advantage of semantic annotations generated from transcription content in video lectures. The recommendation was based on the content covered in different parts of a video lecture. Another contribution was the development of a software system that provides access to the recommendation technique. We evaluated the effectiveness of our technique over a real-world video lecture dataset.

This investigation considered the process of obtaining video transcriptions; the semantic annotation of video content based on available ontology as input; and the recommendation mechanism, which combines the *K-Means* algorithm to cluster video lectures into main sub-areas. In addition, we explored TF-IDF similarity measure which compares semantic annotations from videos and obtains the most similar set of video lectures related to a specific concept. Our solution considered an ontology-based semantic similarity metric to measure how related concepts being approached were according to an ontology. Our evaluation explored a dataset of video lectures from different Computer Science areas to assess the quality of results from our proposed technique in a real-world scenario. We examined the influence of different scenarios of recommendation and how relevant or redundant the recommendations are.

Regarding the recommendation results, our analysis considered three main metrics: MAP@5 (Mean Average Precision), MAR@5 (Mean Average Recall) and Diversity score. Our experiments regarding the recommendation technique showed that it presents good effectiveness. In the under analysis dataset with general MAP for 5 recommendations, we got best results of 77% precision rate. This indicates that the recommendations provided by the system are good to detect similar video lectures over the dataset. The system recommends them in such a way that covers different concepts being approached in different moments of a video.

The low value for the MAR@5 score indicates that our system still has room for improvements. It recommends not only very similar concepts, but concepts that are approached in videos that are related by more specific concepts. For example, in videos where it should recommend concepts related to "Data Structure" sub-area, it recommends broader concepts being approached in a video lecture, such as, "Sorts", "Trees" and other general aspects of the theme. The system can improve to better recognize more specific concepts like: "Selection Sort", "Binary Tree", "Bubble Sort" as important topics the user can be in doubt.

The ontology and the semantic similarity may have a crucial impact on this effect. Further investigations can analyze whether the ontology is good enough to describe the domain; or if the semantic similarity measure should be more important in our similarity equation.

Other interesting point observed in our obtained results was the difference between recommending entire video lectures for a given moment of a video; and recommending just parts of it. This analysis aimed to compare these two approaches to verify which one is further desirable for a video lecture recommendation system. The first scenario obtained a high value for mean average precision and a lower value for the mean average recall. This is valuable for some cases, but cannot be ideal for a learning environment. The second scenario reached a balanced value for Mean average precision and recall, but not so high if compared to the first scenario results.

The effects of this difference can be observed in the Density score in which more different videos were recommended to users. Sometimes these results do not have anything in common. They are recommended because of some content mentioned in some point of the video. For instance, in a video about "Computer Architecture", it mentions the concept of "Java programming Language" and it was recommended to the user as a relevant concept.

We observe some potential improvements in our solution. First, the system performance could better scale as the number of video lectures in the database increase. The issue of this problem concerns the computation of the similarity matrix, which are generated every time a request is made a user. Thus, the system needs to recalculate the similarity over the different videos presented in the database.

The second point is the cold-start problem for new video lectures added in the database. As our technique relies on semantic annotations metadata as the basis for the recommendation, it still suffers from the cold-start problem even though the use of ontology helps to solve this problem. Every new video needs to be semantically annotated and if the ontology does not cover the concepts being approached in the video lecture, any concept is annotated and the system cannot recommend related concepts to the user.

Finally, we consider that our recommendation system meets the requirements for a recommendation and has great potential in helping users in the learning process. We overcome key challenges in this research as our system recommends relevant video lectures for a learner for different moments of a video lecture.

3.6 Conclusion

The use of semantic metadata in video lectures to support recommendation systems are very useful as a way to improve online learning process of the learners and catch different recommendations in the process. In this research, we proposed and designed a semantic-based recommendation system to video lectures relying on the process of extracting relevant concepts from multimedia content based on ontologies. Our developed technique combines different approaches to achieve adequate video recommendations. Our experimental results were relevant to understand the advantages and limitations of our solution. We observed how a semantic approach can improve recommendation in our study scenario, which is the key originality in this investigation. Future work involves the combination of different techniques to achieve further results, such as, use of deep learning models to improve the recommendation results. We plan taking into account information about users in the recommendation process, so we can achieve further personalized recommendations considering not only the learner's doubt, but the context about their learning process and experiences. We also plan to conduct additional user evaluation studies with subjects using our system.

Chapter 4

Discussion

The main research goals of this MS.c. Dissertation were: (1) Study the refinement of concepts in a domain ontology to support semantic annotation of video lecture subtitles. This includes ways of adding new concepts in an existing ontology and adding new vocabulary description to enrich existing concepts in place; (2) Investigate and experimentally evaluate the use of semantic annotation techniques to generate semantic metadata from video lectures; (3) Investigate and evaluate the design of a video lecture recommendation system which explores the extracted semantic annotations; In this Chapter, Section 4.1 discusses how the results generated by the investigations presented in Chapter 3 and Chapter 2 fulfill each of specific objectives of this Dissertation. Section 4.2 presents the limitations observed in the conducted investigations.

4.1 Discussion on the Individual Research Objectives

This section discusses how we fulfill each of specific objectives in the development of this study.

1. **Refine concepts in domain ontology from book resources in natural language.**

   The refinement of concepts in an ontology is a crucial part in the process of improving a domain ontology to support semantic annotation tasks. For achieving this objective, we conceptualized and developed a refinement technique of ontology concepts extracted from book sources to enrich concepts expressed in a specific domain ontology. We explored this concept of refinement from free text book resources in Chapter 2, presenting our technique, which extracts from free text in natural language relevant concepts. We explored information retrieval techniques and inserted new concepts in a domain ontology.

   As we were interested in enriching a domain ontology with new concepts, we explored strategies that enable natural language processing via information retrieval techniques, such as, TF-IDF and lexicon-syntactic analysis from unstructured texts. To this end, we applied our developed technique to different book resources as input in combination with a domain ontology in Computer Science field. Our approach to refine a domain ontology supported the process of video lecture metadata extraction. For this purpose, we refined ontologies in the conceptual level.

In the evaluation of our process, we ran our proposed algorithm to refine a Computer Science Ontology based on the input of book contents from this area. We assessed the effectiveness of our solution in a metadata extraction process with semantic annotation tools. In this context, considering the quality of our refined ontology, we observed that it improved the number and the quality of concepts compared to the initial version of the ontology in the metadata extraction process relying on distinct software tools. We experimentally attested that our proposed technique can refine concepts in an existing domain ontology. The need for manual validation in the insertion of concepts in the ontology and the need for gold standards for comparison purposes are still key challenges.

2. **Evaluate the use of semantic annotation techniques to generate semantic metadata from video lectures.**

   We intensively investigated existing software tools in the state of art that perform semantic annotation as much automatically as possible. We evaluated existing semantic annotation techniques in a metadata extraction process from video lectures. To this end, we considered an evaluation setup where we selected a set of Computer Science video lectures from *Youtube* and explored different ontologies and the selected tools. We conducted this evaluation in Chapter 2. Our study compared not only the effectiveness of the evaluated software tools and ontologies in a metadata extraction process, but the effects of our refined ontology in this process.

   Considering our proposal, we selected a set of Computer Science video lectures from *Youtube* on an extraction process to get the subtitles from these videos. In our study, we used them as input for the semantic annotation tools. The video lectures were selected considering the main sub-areas from the Computer Science, including: Computer Architecture; Data Structure; and Computer Network. In addition, the videos were obtained from *Youtube* randomly, in English and Portuguese language. In this aspect, we assessed the quality of automatic transcriptions associated with the language spoken in the videos, since there is still very little researches exploring Portuguese language in semantic annotations. Concerning the annotations in Portuguese videos, we translated them before the annotations because our refinement and metadata extraction procedures were conducted in the English language. We observed that the use of automatic translation of video subtitles did not negatively influence the annotation process, although our ontology and book sources were in English.

   Based on the results of refinement technique and our analysis of semantic annotation tools, we found that the extraction of semantic metadata in video lectures is a key part to improve the quality and retrieval of videos in recommendation systems. We observed that exploring the best scenario of our refined ontology and semantic annotation tool could improve the effectiveness of recommendations by incorporating the semantic metadata extraction process into the recommendation technique.

3. **Design a video lecture recommendation system based on semantic annotations.**

   Our main objective concerned the developing of an original recommendation system relying on ontology-based semantic metadata automatically extracted. Chapter 2 described our technique to improve the concepts in a domain ontology for this task and evaluated the software tools for semantic annotation of video lecture transcriptions based on available ontologies. Chapter 3 defined our central solution and contribution in this Dissertation, which combined the metadata extraction process with our designed recommendation technique. Our obtained results improved the way in which recommendation of video lectures are provided by recommending content according to ontology concepts appearing in a given part of a video.

   Our recommendation system was designed based on a technique which combines the process of obtaining video transcriptions semantically annotated. The recommendation mechanism uses the *K-Means* algorithm to cluster video lectures in sub-areas. In addition, we applied TF-IDF similarity measure and an ontology-based semantic similarity metric proposed by *Wu and Palmer* [37] to measure how related concepts are according to an ontology. Our evaluation explored a dataset of video lectures from different Computer Science areas to assess the quality of results from our proposed technique in a real-world scenario.

   Concerning our recommendation mechanism, the combination of different techniques was considered to have the best impact in penalizing the diversity of recommendations whereas considering the similarity of related videos, and the semantic aspects of our proposal. *K-Means* was used in our solution to associate videos in the database to main (general) concepts of the case study area. As input for the *K-Means* algorithm, we used annotations from the video lectures in the dataset. All videos were clustered into main groups as initialization for the recommendation engine, in other words, we executed the *K-Means* algorithm one time for the video lecture dataset. In the context of recommendation, TF-IDF and the semantic similarity measure were applied between collections of videos metadata, representing two instances of videos lecture annotations. Provided recommendations were personalized by the concept under analysis (based on the time of the video) considering the semantic and similarity between videos.

   In addition, we examined the influence of different scenarios of recommendation and how relevant or redundant the recommendations were. In these scenarios, we evaluated how our recommendation system behave considering the recommendation of entire videos or only pieces of videos. We found that the provided recommendations were good to detect similar video lectures over the dataset. The system recommends videos in such a way that covers different concepts in different moments of a video. Further investigations concerning the combination of domain ontologies in the recommendation task are needed, additionally the analysis of other similarity metrics to calculate the similarity between different videos are a improvement for this research.

4.2 Limitations

This section describes the limitations of our study.

- The need for manual validation in the conducted evaluations and the quality of the tools to extract text from the book resources are still a limitation of the solution. The extraction of natural language resources from free text are still a key challenge aspect of the area and demand more research.

- In our semantic annotation study, we explored only the use of one ontology per time. This can potentially effect the quality of annotation due to the coverage of the ontology. Further studies could investigate how and to which extent the use of different ontologies in a combined way can improve the metadata extraction. This entails challenges in how to explore more than one ontology in the annotation process.

- Our study did not consider the system performance in a real-world application. Our system could better scale as the number of video lectures in the database increases. In the current version of our solution, the similarity matrix needs to be generated every time a request is made. In this sense, as the number of video lectures in database increases, the calculation of the matrix is affected. This can entail performance issues in our system.

- Our investigation did not study the impacts of the cold-start problem for new video lectures added in the database. Every new video added to the database needs to be semantically annotated. If the concepts of this videos are not expressed in the ontology, no concept is annotated and the system cannot recommend videos associated to related concepts to the user.

- Our obtained results considered a controlled scenario for evaluation purposes. The used videos in the database and the recommendations were supervised by the authors. Additional studies could explore user studies by including students in real-world scenarios.

Chapter 5

Conclusion

The overwhelming number of online video lectures is essential for supporting people's autonomous learning nowadays. However, literature lacks proposals for supporting learners to easily find and correlate contents from online videos. In this MS.c. Dissertation, we assumed that the extraction of semantic metadata from video lectures is essential to improve the recommendation of videos in the learning process support.

This MS.c. Dissertation proposed a semantic-based approach to video lectures recommendation relying on the process of extracting relevant concepts from multimedia content based on ontologies. Our investigation addressed the challenge of extracting fine-grained ontology-based semantic annotations based on the processing of video subtitles in natural language text.

We showed to which extent existing semantic annotation tools provide adequate semantic metadata to be useful for the functioning of semantic-based recommendation systems. We demonstrated that our proposal for refining ontology content influence the results of metadata extraction. Our work defined and implemented a video lecture recommendation system which relies on the metadata extraction process. Our recommendation system explored the semantic annotation for filtering video fragments based on a input concept for searching.

5.1 Summary of Contributions

This MS.c. Dissertation contributes to how people can further explore and take advantage of online video lectures available. We believe our studies have achieved state-of-the-art results in the area of multimedia recommendation. In the following, we summarize the main contributions achieved by the development of this investigation.

1. Our work proposed and implemented an ontology refinement procedure with the aim of turning domain ontology content further specialized (cf. Chapter 2). In particular, we explored book summaries as input data in the Computer Science area to refine a Computer Science Ontology [28]. Our method explored *TF-IDF* technique as a way of detecting those relevant concepts and encode them as adequate subconcepts in the existing ontology. The application of our proposal enabled obtaining a more refined ontology containing additional and relevant concepts. In our

view, applying semantic annotation with more complete domain ontologies enhances metadata extraction results because of the coverage of concepts.

2. We conducted an study for experimentally evaluating the application of several semantic annotation tools for metadata extraction from online video subtitles (cf. Chapter 2). Our study relied on several ontologies and their distinct versions to understand to which extend different tool's approaches and input resources (domain specific ontologies and general ontologies) affect semantic annotation results. We found that distinct ontologies play a key role in the quality of the results in the annotation. The use of ontologies with greater coverage positively influence the metadata extraction. We also found the effects in the annotation quality regarding the input subtitles.

3. We proposed and implemented a semantic-enabled recommendation system suited to filter and search online video lectures based on user's input assigning a doubt flag when interacting with a video (cf. Chapter 3). Our system generates a knowledge base with the annotated videos based on semantic extraction from their subtitles. The developed a recommendation technique that considers the encoded relationships between concepts recognized from the ontology (via the metadata extraction) and cluster related videos. Recommendations are presented based on a ranking basis for end users. In this sense, our solution explored clustering techniques combined with information extraction and similarity computation for calculating semantic similarity between ontology concepts.

4. An online web-based software tool as a prototype that allows learners to watch videos available in our database and obtain the recommendation results (cf. Chapter 3). All the developed code is available in our code repository[1].

5.2 Dissemination

We disseminated our research findings during the development of this MS.c. Dissertation. This work was disseminated via the participation in academic events. We wrote articles published and submitted to international conferences and journals, as follows:

- **BORGES, M. V. M.**; DOS REIS, J. C. and GRIBELER, G. P. 2019. *Empirical Analysis of Semantic Metadata Extraction from Video Lecture Subtitles*. Proceedings of the IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE). Napoli, Italy, pp. 301-306, doi: 10.1109/WETICE.2019.00069.

- **BORGES, M. V. M.** and DOS REIS, J. C. 2019. *Semantic-Enhanced Recommendation of Video Lectures*. Proceedings of the IEEE 19th International Conference on Advanced Learning Technologies (ICALT). Maceio, Brazil, pp. 42-46, doi: 10.1109/ICALT.2019.00013.

[1]https://gitlab.ic.unicamp.br/jreis/semantic-metadata-extraction

- **BORGES, M. V. M.** and DOS REIS, J. C. *Ontology Refinement for Supporting Semantic Metadata Extraction from Video Lecture Subtitles.* Submitted to an international journal.

- **BORGES, M. V. M.** and DOS REIS, J. C. *Video Lecture Recommendations based on Semantic Annotations.* Submitted to an international journal.

5.3 Future Work

This section presents perspectives of future investigations to expand the studies addressed in this MS.c. Dissertation:

- Further evaluate the development of techniques to enrich ontology content from book resources. For example, explore other input resources (*e.g.*, Web page contents) and other existing types of techniques for ontology learning;

- Investigate how the automatic extraction of ontology relations in the ontology refinement process can leverage metadata extraction techniques;

- Extensively experiment other distinct software annotation tools and ontologies in the metadata extraction process;

- Investigate how to capture and consider learners' preferences and styles as input for the recommendation system;

- Conduct a pilot study with human learners using our developed systems to qualitatively evaluate the recommendation results;

- Conduct a usability study in our implemented software system to further improve user experience with the software tool.

Bibliography

[1] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge & Data Engineering*, 17(6):734–749, 2005.

[2] Patrick Arnold and Erhard Rahm. Extracting semantic concept relations from wikipedia. In *Proceedings of the 4th International Conference on Web Intelligence, Mining and Semantics (WIMS14)*, pages 1–11, 2014.

[3] Muhammad Nabeel Asim, Muhammad Wasim, Muhammad Usman Ghani Khan, Waqar Mahmood, and Hafiza Mahnoor Abbasi. A survey of ontology learning techniques and applications. *Database*, 2018(1):bay101, 2018.

[4] Marcos Vinicius Macedo Borges, Julio Cesar dos Reis, and Guilherme Pereira Gribeler. Empirical analysis of semantic metadata extraction from video lecture subtitles. In *2019 IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, pages 301–306. IEEE, 2019.

[5] Lucas F Brunialti, V Freire, SM Peres, and Clodoaldo AM Lima. aprendizado de máquina em sistemas de recomendação baseados em conteudo textual uma revisao sistematica. In *XI Brazilian Symposium on Information System*, pages 203–210, 2015.

[6] Sílvio César Cazella, MASN Nunes, and Eliseo Reategui. A ciência da opinião: Estado da arte em sistemas de recomendação. *André Ponce de Leon F. de Carvalho; Tomasz Kowaltowski..(Org.). Jornada de Atualização de Informática-JAI*, pages 161–216, 2010.

[7] Li Chen, Wen Wu, and Liang He. Personality and recommendation diversity. In *Emotions and Personality in Personalized Services*, pages 201–225. Springer, 2016.

[8] Massimiliano Ciaramita, Aldo Gangemi, Esther Ratsch, Jasmin Saric, and Isabel Rojas. Unsupervised learning of semantic relations between concepts of a molecular biology ontology. In *IJCAI*, pages 659–664, 2005.

[9] Sandro Athaide Coelho and Jairo Francisco de Souza. Anotação semântica de transcritos para indexação e busca de vídeos. In *Conferência Ibero Americana (WWW/INTERNET).[GS Search]*, 2014.

[10] Hadi Khosravi Farsani and Mohammadali Nematbakhsh. A semantic recommendation procedure for electronic product catalog. *International Journal of Applied Mathematics and Computer Sciences*, 3(2):86–91, 2006.

[11] Dieter Fensel. Ontologies. In *Ontologies*, pages 11–18. Springer, 2001.

[12] Mingxin Gan, Xue Dou, and Rui Jiang. From ontology to semantic similarity: calculation of ontology-based semantic similarity. *The Scientific World Journal*, 2013, 2013.

[13] Thomas R Gruber. Toward principles for the design of ontologies used for knowledge sharing? *International journal of human-computer studies*, 43(5-6):907–928, 1995.

[14] Greg Hamerly and Charles Elkan. Learning the k in k-means. *Advances in neural information processing systems*, 16:281–288, 2004.

[15] Andrew R Hippisley, David Cheng, and Khurshid Ahmad. The head-modifier principle and multilingual term extraction. *Natural Language Engineering*, 11(2):129, 2005.

[16] Sungrim Kim and Joonhee Kwon. Effective context-aware recommendation on the semantic web. *International Journal of Computer Science and Network Security*, 7(8):154–159, 2007.

[17] Atanas Kiryakov, Borislav Popov, Ivan Terziev, Dimitar Manov, and Damyan Ognyanoff. Semantic annotation, indexing, and retrieval. *Web Semantics*, 2(1):49–79, 2004.

[18] Matevž Kunaver and Tomaž Požrl. Diversity in recommender systems–a survey. *Knowledge-based systems*, 123:154–162, 2017.

[19] Jiaqing Liang, Yi Zhang, Yanghua Xiao, Haixun Wang, Wei Wang, and Pinpin Zhu. On the transitivity of hypernym-hyponym relations in data-driven lexical taxonomies. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

[20] Antonis Loizou and Srinandan Dasmahapatra. Recommender systems for the semantic web. In *ECAI 2006 Recommender Systems Workshop*, 2006.

[21] Lluís Màrquez, Lluis Padro, and Horacio Rodríguez. A machine learning approach to pos tagging. *Machine Learning*, 39(1):59–91, 2000.

[22] Lailil Muflikhah and Baharum Baharudin. Document clustering using concept space and cosine similarity measurement. In *2009 International conference on computer technology and development*, volume 1, pages 58–62. IEEE, 2009.

[23] Natalya F Noy, Nigam H Shah, Patricia L Whetzel, Benjamin Dai, Michael Dorf, Nicholas Griffith, Clement Jonquet, Daniel L Rubin, Margaret-Anne Storey, Christopher G Chute, et al. Bioportal: ontologies and integrated data resources at the click of a mouse. *Nucleic acids research*, 37(suppl_2):W170–W173, 2009.

[24] Eyal Oren, Knud Möller, Simon Scerri, Siegfried Handschuh, and Michael Sintek. What are semantic annotations. *DERI Galway*, 9:62, 2006.

[25] Eduardo Peis, JM Morales del Castillo, and Juan A Delgado-López. Semantic recommender systems. analysis of the state of the topic. *Hipertext. net*, 6(2008):1–5, 2008.

[26] Anand Rajaraman and Jeffrey David Ullman. *Mining of massive datasets*. Cambridge University Press, 2011.

[27] Lev Ratinov and Dan Roth. Design challenges and misconceptions in named entity recognition. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*, pages 147–155. Association for Computational Linguistics, 2009.

[28] Angelo A Salatino, Thiviyan Thanapalasingam, Andrea Mannocci, Francesco Osborne, and Enrico Motta. The computer science ontology: a large-scale taxonomy of research areas. In *International Semantic Web Conference*, pages 187–205. Springer, 2018.

[29] Gerard Salton and Christopher Buckley. Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5):513–523, 1988.

[30] David Sánchez and Antonio Moreno. Learning non-taxonomic relationships from web documents for domain ontology construction. *Data & Knowledge Engineering*, 64(3):600–623, 2008.

[31] Celso AS Santos, Alexandre Santos, and Tatiana A Tavares. Uma estratégia para a construção de ambientes para a descrição semântica de vídeos. In *Proc of*, pages 274–281, 2007.

[32] Marcio de Carvalho Saraiva and Claudia Bauzer Medeiros. Correlating educational documents from different sources through graphs and taxonomies. In *Brazilian Symposium on Databases*, 2018.

[33] Nigel Shadbolt, Tim Berners-Lee, and Wendy Hall. The semantic web revisited. *IEEE intelligent systems*, 21(3):96–101, 2006.

[34] Thabet Slimani. Semantic annotation: The mainstay of semantic web. *International Journal of Computer Applications Technology and Research*, abs/1312.4794, 2013.

[35] Mari Aulikki Virtanen, Elina Haavisto, Eeva Liikanen, and Maria Kääriäinen. Ubiquitous learning environments in higher education: A scoping literature review. *Education and Information Technologies*, 23(2):985–998, 2018.

[36] Rui-Qin Wang and Fan-Sheng Kong. Semantic-enhanced personalized recommender system. In *Machine Learning and Cybernetics, 2007 International Conference on*, volume 7, pages 4069–4074. IEEE, 2007.

[37] Zhibiao Wu and Martha Palmer. Verb semantics and lexical selection. *arXiv preprint cmp-lg/9406033*, 1994.

[38] Zhiwen Yu, Yuichi Nakamura, Seiie Jang, Shoji Kajita, and Kenji Mase. Ontology-based semantic recommendation for context-aware e-learning. In *International Conference on Ubiquitous Intelligence and Computing*, pages 898–907. Springer, 2007.